DOCUMENT RESUME

ED 331 296 FL 019 134

AUTHOR Gutstein, Shelley; Goodwin, Sarah H.
TITLE The CLEAR Oral Proficiency Exam (COPE) Project Report and Addendum: Clinical Testing and Validity and Dimensionality Studies.
INSTITUTION Center for Applied Linguistics, Washington, D.C.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 16 Dec 87
CONTRACT 400-85-1010
NOTE 55p.; The Addendum Report, 1988, was prepared by Lih-Shing Wang, Gina Richardson, and Nancy Rhodes.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Elementary Secondary Education; English (Second Language); *Immersion Programs; *Language Proficiency; *Language Tests; Spanish; *Test Construction; *Verbal Tests
IDENTIFIERS *CLEAR Oral Proficiency Exam

ABSTRACT

The process of developing, piloting, and refining an oral proficiency test as part of the Center for Language Education and Research (CLEAR) test battery is described. The test was designed to fill a need for an oral interview-type measure adapted for fifth to seventh grade students. The test was produced first for Spanish immersion students, then adapted for English-as-a-Second-Language (ESL) students at the same grade level, resulting in the CLEAR Oral Language Proficiency Exam (COPE) in Spanish and English. The report has five sections. The first discusses the background of the project briefly. The second outlines precedents and procedures. Sections three and four describe the COPE-Spanish trial administrations and ESL adaptation and trials, respectively. The final section offers conclusions and recommendations for this segment of test development. Clinical testing and validity and dimensionality recommended studies for the exam are reported in a 1988 addendum report. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED331296

THE CLEAR ORAL PROFICIENCY EXAM (COPE)

Project Report by

Shelley Gutstein, Ph.D. and Sarah H. Goodwin, Ed.D.

for the Center for Applied Linguistics

Washington, D. C.

December 16, 1987

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Tucker, G.R.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CAL

Center for Applied Linguistics

FL 019134

CONTENTS

I.	Background of the Project	Page 1
II.	Precedents and Procedures	1
III.	COPE-Spanish Trial Administrations	6
IV.	COPE-ESL Adaptation and Try-Outs	10
V.	Conclusions and Recommendations	16

Clear Oral Proficiency Exam (COPE)

Project Report

I. Background of the Project

The CLEAR research studies on second language instruction proposed in the year two work plan[1] identify as a major need the development of oral proficiency tests of Spanish and English (ESL). The work plan reviewed available test instruments covering cognitive-academic language as well as social communication skills and found that suitable instruments to evaluate oral proficiency at the upper elementary school levels did not exist. The "CLEAR Test Battery" will include the IDEA Proficiency Test (IPT) and the Woodcock Language Proficiency Battery, which are discrete-point tests and do not produce conversational speech samples. Thus the need was identified for an oral interview-type test adapted for fifth to seventh grade students which would elicit normal speech and give global scores on a rating scale. The mandate for the test developers was to produce first an oral proficiency test for fifth to seventh grade students in Spanish immersion programs, then adapt it for English as a second language (ESL) students of the same grades. The resulting test would be called the CLEAR Oral Proficiency Exam (COPE) with Spanish and English versions.

II. Precedents and Procedures

A review of recent studies on oral proficiency testing was carried out and provided the following background information. An oral proficiency test must be distinguished from an oral achievement test in that it "compares the student's speaking

ability with that of a well-educated native speaker using the language for real-life communicative purposes as contrasted with an [oral] achievement test, which is based on material covered in a particular course of study."[2]

The prototype oral interview proficiency test is the one developed in 1956 by the Foreign Service Institute of the U. S. State Department and used since then with some minor modifications by U. S. government agencies and some schools and colleges.[3] In this test each subject is interviewed by one or two trained testers who ask a series of open-ended questions designed to elicit by the end of the interview the highest level of speech competency. The resulting speech sample is scored on a scale of 0 to 5, based on explicit descriptions of comprehension, fluency, vocabulary, grammar, and pronunciation at each level. There have been a number of adaptations of the FSI format for college and secondary use, the most recent of which is the descriptive rating scale produced by the American Council on the Teaching of Foreign Languages and the Educational Testing Service (ETS), known as the "ACTFL Proficiency Guidelines."[4] These guidelines expand somewhat the lower levels of the FSI scale and use descriptive terms in place of numbers to describe each level.

The ACTFL/ETS rating scale was adapted by Educational Testing Service researchers in 1983 for junior high school students in French immersion programs in New Brunswick, Canada. The major modifications were adjustments in the descriptors for grammar and omission of pronunciation, since this aspect of language ability, the researchers found, "ceases to be a matter

of much concern with students exposed to [a foreign language] from the early grades."[5] Since this modified ACTFL/ETS rating scale was based on interviews with immersion program students only slightly older than the CLEAR target audience, it seemed appropriate to use it as a starting point for developing an oral proficiency test for fifth to seventh grade students in Spanish immersion program in the U. S.

In addition to the general requirement for an oral proficiency test as opposed to an oral achievement test, other specifications for the COPE were:

1. The test would require 15-20 minutes for a single administration.
2. It could be administered by school teachers or principals without special/extensive training.
3. It would overcome the boredom produced by the question-answer format of some oral interviews.
4. The test would assess cognitive/academic language, as well as social/survival language.

To develop a test meeting these specifications the following procedures were carried out.

1. In order to secure background information on classroom/academic language, the test developer observed French and Spanish immersion classes at Oak View Elementary School in Montgomery County, Md., and surveyed social studies and science textbooks (Spanish) used in Maryland and other immersion programs.

2. To reduce the overall time required for testing and to deal with the boredom/interest factor, the interview format was modified in two ways:

   --two students at a time are interviewed.

   --the entire interview is "contextualized."

   Omaggio's[6] discussion of teaching and testing language in context suggested the value of "contextualizing" the testing situation for COPE. Also, in discussing various interview formats, Omaggio describes paired interviews with two students taking turns asking questions based on conversation cards and responding. Another precedent for interviewing two rather than only one student at a time is found in Reschke's[7] suggested modifications of the FSI oral interview for secondary and college students. He proposed interviewing as many as three to five students at a time. We found that keeping track of two students at a time was as much as one interviewer could handle.

3. The format for the interview thus became creating an imaginary, but realistic, situation in which two students carry out a series of brief conversations based on instructions contained in a set of dialogue cards. In other words, for COPE--Spanish, students are asked to play the roles of a Mexican student visiting a U. S. school and a North American student acting as a guide during the visit, developing brief conversations

based on instructions found in three to five of the dialogue cards.

4. Contents of the dialogue cards were developed on the basis of observations of school and classroom activities, interviews with sixth grade students to find out their interests, textbook contents, and several ideas from a 1984 set of ETS Situation Cards.[8] There were 20 cards in the set as first designed: two pink cards (introductory) used with all administrations of the test, nine green cards (beginning/intermediate), and 11 orange cards (intermediate/advanced). Instructions to the students were given in the third person.

5. In order to make scoring easier, the descriptions of the modified ACTFL/ETS rating scale were organized at each level by the categories of comprehension, fluency, vocabulary, and grammar (omitting references to French) and put into a matrix. The concept of this type of matrix and the instructions for scoring are based on the "Student Oral Proficiency Rating (SOPR)."[9] The more technical terminology of the modified ACTFL/ETS scale was simplified, and where it lacked descriptors specific to a category level, language was used from the SOPR or from a much- abbreviated version of the ACTFL/ETS guidelines prepared for elementary schools.[10]

Thus, before the try-outs, procedure for the Clear Oral Proficiency Exam (COPE)--Spanish provided for:

--group instructions in Spanish which explained the test and described the context for the paired interviews--a visit of a Mexican student to a U. S. elementary school with a Spanish immersion program.

--a 15-20 minute session with each pair of students who were asked to respond to cues in Spanish read by the tester from four to six of the 20 dialogue cards (tape recorded if at all possible.)

--rating of the resulting language samples by indicating on a nine-level, four-category rating scale which level best described the sample in the categories of comprehension, fluency, vocabulary, and grammar by the test administrator at the time of testing.

III. COPE-Spanish Trial Administrations

The purpose of the trials was to test the mechanics of the test construction: the ease with which the test could be administered and the ability of individual test items to generate an oral language sample large and complete enough to be evaluated.

The COPE-Spanish tryouts were held at three locations in late May/early June, 1987. A total of twenty-seven children were tested at three sites: Culver City, California (four subjects), Milwaukee, Wisconsin (thirteen subjects) and Silver Spring, Maryland (ten subjects). The three trials all followed the same format of instructions, situation cards and paired interviews, as described above. There were, however, some important differences in the trials, and these are summarized in Table 1 below.

First, the Maryland try-out was conducted with students in a partial immersion program, while the California and Wisconsin try-outs were conducted with students in total immersion programs. Not surprisingly, the range of scores (evaluations) for the Maryland subjects is somewhat lower than that of the other two groups.

A second important difference is that the number of raters varied with each trial. In the Maryland test, two raters were in the testing room with each pair of subjects. One rater conducted each interview while the other observed and made notes. The raters alternated interviewing, and both evaluated all the subjects. In California, one rater worked with each pair of subjects. For the Wisconsin test, three independent raters worked with each pair of subjects. The interviews in California and Maryland were tape recorded.

There were also some differences in the ways the various raters used the situation cards during the tryouts. The rater in California used the same situation cards with each pair of students. In the Maryland try-out, a different assortment of green and orange cards was used with each pair in an attempt to obtain feedback on content and usability of all twenty situations.

After the trials were completed, the Wisconsin and California raters were debriefed by telephone to obtain reactions to the instrument, as well as suggestions and comments regarding revisions and corrections. Both sets of raters also provided some written feedback. (See Appendix 1 for copies of their

comments.) Based on this feedback and their own observations, the test developers made the following changes in the COPE-Spanish.

1. The number of situation cards was reduced from twenty to seventeen, without changing the three original levels of difficulty.
2. The language of the situation cards was revised, simplified, and to the extent possible, placed in second person address.
3. The content of the situation cards was revised to ensure equal oral production on the part of both interlocutors in the dialogue.
4. The group instructions were simplified. It was decided that these instructions should be presented to the subjects in English in order to ensure comprehension. Group oral instructions would be given in English and the remainder of the instructions and conversation would be given in Spanish.
5. The number of repetitions of each situation card was limited to two, and prompting of the students was to be discouraged.
6. The rating scale matrix was found to be usable and workable, requiring only a few minor clarifications and rewordings. Some of the descriptions of skills/levels were shortened and made more concise.

Several research questions emerged from the COPE-Spanish try-outs. These are discussed below in the final section, Conclusions and Recommendations.

TABLE 1 DIFFERENCES IN THE COPE-SPANISH TRIALS

	MARYLAND	CALIFORNIA	WISCONSIN
Type of Program	Partial Immersion	Total Immersion	Total Immersion
Testing Situation	Two raters w/ each pair	One rater w/ each pair	Three indep. raters
Time	Approx. 20 min./pair	Not monitored	Approx. 20 min./pair
N=	10	4	13
Situation Cards	Varied w/ pair	Same cards w/ pair	Varied w/ pair
Pre-test Proficiency Level	All levels	Middle level Screened by teacher	All levels
Range of Rating	Novice low- Jr. Inter. Mid	Jr. Inter. Mid- Jr. Inter. High	Jr. Inter. Low- Superior
Tapes of Try-outs	Available	Available	Not avail.

IV. ESL Adaptation and Try-Outs

Adapting the COPE-Spanish for ESL students required in the first stage (1) translating the Spanish dialogue cards into English and (2) modifying the grammatical descriptions in the rating scale to reflect English morphology and syntax. Although the group instructions remained in English, this represented a change in administration procedure from the final form of the COPE-Spanish, since the COPE-ESL instructions would not be given in the students' native language. Translations into all mother tongues were not possible.

The trials for the COPE-ESL were conducted on July 23, 1987 at Sleepy Hollow Elementary School in Fairfax County, Virginia. This school was the site of a summer intensive ESL program for elementary school students who were bussed from various areas of the county to participate in the program. Students received three hours of ESL instruction daily.

The twenty subjects for the COPE-ESL trials were selected from two classrooms. The first classroom consisted of students who would be entering grades five and six in the fall, and whose English proficiency was considered weak, following Fairfax County placement and achievement criteria. Ten students were selected from this classroom. The second classroom consisted of students who would be entering grades five, six, and seven in the fall, and whose English proficiency was considered stronger. Many of these students had been in the United States longer than the students in the first classroom. Ten students were selected from this classroom as well. The twenty students included four fifth

graders, 13 sixth graders, and two seventh graders. They represented seven language backgrounds, with Spanish being the most frequent (seven subjects), followed by Korean (four subjects). Students ranged in age from ten to fourteen years old. Tables 2 and 3 contain data for the subjects in the trials which were obtained directly from the subjects.

The teacher in each classroom selected the students to participate in the trials. Teachers were told to choose students who represented the entire range of ESL proficiency. No other criteria were given to the teachers to use in making their selections.

Each tester worked with one of the classroom groups. The trials were conducted in two quiet rooms provided by the school. After "setting up", each tester returned to the assigned classroom, called the selected students out, and, following the test instructions, read the preliminary instructions to the students. The trials were then conducted, following the original dialogue card format. Each pair of students was tested for twenty minutes, and the testing sessions were tape recorded.

RESULTS FROM CLASSROOM #1: The tester was unable to follow the original format because (1) the majority of the students did not comprehend the testers oral instructions and cues, and (2) the students had difficulty understanding each other, and were unable to enter into the role-play situations. (Possibly there were other factors which caused the lack of success with this format.) The tester, therefore, elected to use those dialogue cards which were easily adapted to the one-on-one style of

testing. She was able to elicit language from many of the students in this manner.

RESULTS FROM CLASSROOM #2: The tester was able to follow the intended format fairly closely. Three of the five pairs of students were able to role play easily in all situations, one pair had difficulty with the role-play format, and the last pair, which consisted of two very shy students, did not role play in any of the situations. The tester used different situation cards with each pair of students, in an attempt to try out all of the cards.

The results of the COPE-ESL trials indicated that several modifications in format and content were necessary.

1. Pronunciation was determined to be an important variable in intelligibility, in contrast to the findings of Rabiteau and Taft.[11] It was therefore decided to add a series of evaluation categories for pronunciation to the evaluation grid.
2. During the trials some of the students had difficulty with the content material in some of the situation cards. Later consultations with their teachers revealed that students at this grade level were not receiving instruction in the content areas of science and social studies. For this reason, we revised several of the situations and eliminated those which contained vocabulary and concepts not yet studied.
3. Since the COPE-ESL was intended to be used with any fifth to seventh grade student, it was necessary to modify the test

to accommodate students at the low end of the proficiency scale. Results of the COPE-ESL trials indicated that the situation-based role-play format was not be an effective language elicitation device with students of very low English proficiency. Some students did not have the linguistic flexibility to cope with a hypothetical situation in English. Cultural differences might also have affected their ability to deal with role plays, and many of the students could not understand the instructions. It became clear that language samples simply could not be elicited from these students using the dialogue format.

It was therefore decided to revise the test to include a fourth component: one-on-one questions. The purpose of these questions, inserted between the warmup and the role plays, is to provide the tester with a language sample sufficient to determine whether or not the students are capable of participating in the role play situations. Based on the student's performance on the one-on-one questions, the tester decides whether to go on to the role plays presented in the dialogue cards. The resulting format then consisted of the warmup, one-on-one questions, role plays and wind down. The same evaluation grid is used with the revised version of the test. Further information regarding administration of the revised test can be found in the instructions for COPE-Spanish.

TABLE 2 STUDENT DATA CLASSROOM 1

STUDENT	GRADE	AGE	COUNTRY	LANGUAGE	LENGTH OF TIME FFX. SCHOOLS	YRS. SCHOOL HOME COUNTRY	YEARS OF ENGLISH HOME COUNTRY
1	6	10	Pakistan	Urdu	3 Years	6 years	None
2	6	12	Korea	Korean	1 Year	6 years	None
3	5	12	Vietnam	Vietnamese	2 Years	None	None
4	5	10	Nicaragua	Spanish	8 Months	5 years	None
5	6	12	Nicaragua	Spanish	1 Year	2 years	None
6	6	12	El Salvador	Spanish	5 Months	6 years	None
7	7	12	Korea	Korean	5 Months	6 years	None
8	6	10	Argentina	Spanish	5 Months	7 years	None
9	6	12	Palestine	Arabic	8 Months	8 years	4 years
10	6	12	Korea	Korean	1 Year	5 years	None

TABLE 3 STUDENT DATA CLASSROOM 2

STUDENT	GRADE	AGE	COUNTRY	LANGUAGE	LENGTH OF TIME FFX. SCHOOLS	YRS OF SCHOOL HOME COUNTRY	YRS OF ENGLISH HOME COUNTRY
1	6	11	Korea	Korean	15 Months	4 Years	None
2	5	11	Cambodia	Cambodian (Thai, Lao)	7 Years	?	None
3	6	11	El Salvador	Spanish	1 1/2 Years	5 1/2 Years	None
4	7	13	Pakistan	Urdu, Russian, Serbo-Croatian	1 Year	7 Years	3 Years
5	6	14	Cambodia	Cambodian	4 Years	1 Year	None
6	6	10	India	Hindi	1 Year	5 Years	4 Years
7	5	9	Peru	Spanish	1 1/2 Years	3 Years	3 Years
8	7	14	Vietnam	Vietnamese	2 Years	6 Years	6 Years
9	6	10	Palestine	Arabic	2 Years	4 Years	4 Years
10	6	11	Guatemala	Spanish	3 Years	2 Years	None

V. Conclusions and Recommendations

On the basis of these limited trials, there is considerable evidence that the COPE in both Spanish and ESL versions meets the original specifications as outlined on page 3. However, the following additional work is recommended before extensive use is made of the COPEs.

1. Full-blown field tests with larger numbers of subjects are needed to evaluate reliability and validity.
2. More study is needed to establish the relative difficulty of the various dialogue cards so that they can be ordered according to the level of difficulty.
3. The Spanish dialogue cards should be reviewed by a native speaker of Latin American Spanish who is experienced in testing.
4. Language functions should be studied in terms of language actually produced as well as according to the functions suggested by the dialogue cards.
5. More study is needed to determine the criteria for moving from one-to-one questions to role plays in the ESL test.
6. It may be necessary to deal further with the implications for testing of the differences in the two populations for which the COPE-Spanish and COPE-ESL are designed. The American immersion program students and the primarily immigrant students in U. S. ESL programs differ greatly, not only in previous education, cultural and language backgrounds, but also in the

length of exposure to the target languages and methodologies of the instruction received in those languages.

NOTES

1. Submitted to the Office of Educational Research and Improvement, U. S. Department of Education. Center for Language Education and Research. Year Two Workplan. Task Six: Second Language Instructional Program.

2. Pardee Lowe Jr. and Judith E. Liskin-Gasparro, Testing Speaking Proficiency: The Oral Interview (ERIC Clearinghouse on Languages and Linguistics, Washington, D. C., 1986) p. 1.

3. Claus Reschke, Adaptation of the FSI Interview Scale for Secondary Schools and Colleges. In Clarke, John L. D., ed., Direct Testing of Speaking Proficiency: Theory and Application: Proceedings of a Two-Day Conference. (Educational Testing Service, Princeton, 1978) pp. 77, 78

4. Alice C. Omaggio, Teaching Language in Context (Heinle and Heinle Publishers, Inc. Boston, 1983) pp. 433-443. These guidelines are reproduced and discussed in other publications as well.

5. Kathleen Rabiteau and Hessy Taft, Provisional Modified ACTFL/ETS Oral Proficiency Scale for Junior High School Students (Educational Testing Service, Princeton, N. J. n.d., mimeographed) p. 3.

6. Omaggio, op. cit., and Proficiency-Oriented Classroom Testing (Center for Applied Linguistics, Washington, D. C., 1983)

7. Reschke, op. cit., p. 81.

8. Educational Testing Service, Situation Cards: Designed and assembled with the assistance of the Interagency Language Roundtable participating agencies. 1984.

9. Annette M. Zehler, Inservice Session on the Student Oral Proficiency Rating (SOPR) (Development Associates, Inc., Arlington, Va., 1987)

10. B. Harris-Schenz and B. Hicks, ACTFL/ETS Oral Proficiency Guidelines (Modified for use at East Hills, Liberty, and Linden Elementary Schools, 1986)

11. Rabiteau and Taft, op. cit., p.3.

APPENDIX 1

SUGGESTED REALIA FOR USE WITH THE COFE-SPANISH AND COFE-ESL

Time line

Map of the United States

Drawings:

- Library
- School Bus
- Fire Drill (Children lined up)
- Children on Playground
- Scientific Equipment
- TV, Records, Clothes
- Outside of a movie theater
- An Automobile Accident

BEST COPY AVAILABLE

APPENDIX 2

LETTERS AND NOTES FROM CALIFORNIA AND WISCONSIN TRYOUTS

BEST COPY AVAILABLE

milwaukee public schools

SPANISH IMMERSION PROGRAM
FIFTY-FIFTH STREET SCHOOL
2765 S. Fifty-Fifth Street
Milwaukee, Wisconsin 53219-3269
Area 414: 327-5780

June 17, 1987

Ms. Sarah Goodwin
5624 N. 5th Street
Arlington, Virginia 22205

Dear Sarah:

Enclosed you will find the rating sheets for the 13 students we tested. As I explained on the phone three different people tested the youngsters because of the lack of time and the duties each of us had. We all agree that it would be better if one person did all the testing because all would be more similarly assessed.

The Northamerican student does considerably more talking. It should be equally divided for each diálogo. For example, in #3 have the Mexican student explain how lunch is handled at his school.

Most students did not want the dialogue explanation read twice as they understood it the first time. However, they would forget some details they were to include in their discussion. Perhaps the student could have the card, but may look at it only as a reminder of a detail.

Neither "prop" was particularly useful. The students had no difficulty with a schedule and the map was so dark they could not refer to it.

Often times we wanted to take a part of two ratings and put them together rather than "X" just one. For that reason we felt a checklist would be better.

We are pleased you have devised an instrument as there is definitely a need for one. We will be happy to cooperate with you in this regard in the future.

Sincerely,

J Hochstatter

Jeanne Hochstatter
Principal

JH/cg

P.S. I did make certain word changes on the cards and on page 2 of the guide. I am returning them as it may be useful to you.

encl

BEST COPY AVAILABLE

Mary Hawkins

OBSERVATIONS

Role playing

- The students enjoyed the experience although they were a bit too concerned with the prompts I had given them. They hesitated about ad libbing.

- Unclear in the directions about how much explaining the prompter can do.

- rate sheet is difficult A) to read b) to decide which category to put the student in.

- categories were good

BEST COPY AVAILABLE

APPENDIX 3

LANGUAGE FUNCTIONS IN THE COPE-SPANISH AND COPE-ES_

(Note: This analysis of the language functions of the dialogue cards is based on the attached list of functions from Omaggio, Teaching Language in Context.)

BEST COPY AVAILABLE

COPE-SPANISH

SITUATION CARDS

Language Functions--Preliminary List

Situation Number	Title	Language Functions
1	Presentactiones	Informal greetings 6.1 Getting to know each other 6.3 Introducing oneself 6.7
2	El programa de estudios	Getting to know each other 6.3 Expressing liking 3.1 Stating Factual Information 1.2
3	La cafeteria	Asking/Receiving Information 1.7 Expressing Liking/Disliking 3.1 Stating Factual Information 1.1 Stating Factual Information 1.1 Describing 2.44
4	Lineas cronologicas	Reporting 1.6 Stating Factual Information 1.2 Describing/Narrating 2.44
5	La biblioteca	Asking/Receiving Information 1.7 Reporting 1.6 Explaining how something works 1.9
6	Practica de incendios	Explaining how something works 1.9 Expressing fear/worry 3.6
7	Dos viajes	Stating want/desire 3.16 Proposing a course of action 5.13 Reporting 1.6
8	Autobuses escolares	Explaining how something works 1.9 Asking/Receiving Information 1.7
9	Al Cine	Extending invitation/offer 2.5 Accepting/Declining Inv./offer 2.6
10	La vida social	Inquiring about belief/opinion 2.36 Expressing belief/opinion 2.35 Describing/narrating 2.44 Gossiping/Telling Secrets 6.24
11	Una Fiesta	Describing/Narrating 2.44 Asking/Receiving Info. 1.7 Passing on information 7.13

BEST COPY AVAILABLE

Situation Number	Title	Language Functions
12	Proyecto de ciencias	Explaining how something works 1.4 Asking about or seeking factual information 1.3
13	Carreras futuras	Expressing possibility 2.13 Stating want/desire 3.16
14	Un choque	Expressing fear/worry 3.6 Describing (an event) 2.44 Asking for/Receiving Info. 1.7
15	Una pelea	Asking for/Receiving Info. 1.7 Describing 2.44
16	Reglas injustas	Inquiring about belief/opinion 2.35 Expressing belief/opinion 2.34 Describing 2.44
17	Equipo cientifico	Describing 2.44 Asking about/seeking factual info 1.3

BEST COPY AVAILABLE

COPE-ESL

SITUATION CARDS

Language Functions--Preliminary List

ONE-TO-ONE QUESTIONING

STUDENT A:	Time Line	Reporting 1.6 Stating Factual Infor. 1.2 Describing/Narrating 2.44
STUDENT B:	Family	Reporting 1.6 Stating Factual Infor. 1.2 Describing/Narrating 2.44
BOTH:	La cafeteria	Asking/Receiving Information 1.7 Expressing Liking/Disliking 3.1 Stating Factual Information 1.1 Stating Factual Information 1.1 Describing 2.44

PAIRED CONVERSATIONS

Situation Number	Title	Language Functions
1	School buses	Explaining how something works 1.9 Asking/Receiving Information 1.7
2	The library	Asking/Receiving Information 1.7 Reporting 1.6 Explaining how something works 1.9
3	Fire drill	Explaining how something works 1.9 Expressing fear/worry 3.6
4	The movies	Extending invitation/offer 2.5 Accepting/Declining Inv./offer 2.6
5	Social life	Inquiring about belief/opinion 2.36 Expressing belief/opinion 2.35 Describing/narrating 2.44 Gossiping/Telling Secrets 6.24
6	A party	Describing/Narrating 2.44 Asking/Receiving Info. 1.7 Passing on information 7.13

BEST COPY AVAILABLE

Situation Number	Title	Language Functions
7	Future careers	Expressing possibility 2.13 Stating want/desire 3.16
8	An Automobile Accident	Expressing fear/worry 3.6 Describing (an event) 2.44 Asking for/Receiving Info. 1.7
9	A fight	Asking for/Receiving Info. 1.7 Describing 2.44
10	Scientific equipment	Describing 2.44 Asking about/seeking factual info 1.3

BEST COPY AVAILABLE

Appendix B

Greek Basic Course Functions Catalog: Specific List of Contents

F1 Imparting and Seeking Factual Information

1.1 Identifying Objects, Persons, Processes
1.2 Stating Factual Information
1.3 Asking About or Seeking Factual Information
1.4 Stating Hypothesis
1.5 Stating Generalization
1.6 Reporting
1.7 Asking/Receiving Information
1.8 Summarizing
1.9 Explaining How Something Works

F2 Expressing and Determining Intellectual Attitudes

2.1 Expressing Agreement and Disagreement
2.2 Inquiring About Agreement/Disagreement
2.3 Expressing Understanding/Failure to Understand
2.4 Admitting (Affirming/Denying)
2.5 Extending Invitation/Offer
2.6 Accepting/Declining Invitation/Offer
2.7 Inquiring Whether Invitation Is Accepted or Declined
2.8 Offering to Do Something
2.9 Stating Intentions
2.10 Inquiring About Intention(s)
2.11 Stating Warning
2.12 Inquiring About Remembering/Forgetting
2.13 Expressing Possibility/Impossibility
2.14 Inquiring Whether Something Is Impossible/Possible
2.15 Expressing Capability/Incapability
2.16 Inquiring About Capability/Incapability
2.17 Expressing Need
2.18 Inquiring About Need

2.19 Expressing Certainty/Uncertainty
2.20 Inquiring About Certainty/Uncertainty
2.21 Expressing Obligation/Non-Obligation
2.22 Inquiring About Obligation/Non-Obligation
2.23 Granting/Withholding Permission
2.24 Requesting Permission
2.25 Asking if Others Have Permission
2.26 Stating That Permission Is Withheld
2.27 Expressing Confirmation
2.28 Confirming a Known Fact
2.29 Inquiring About Denial
2.30 Expressing That a Speaker Expects a Positive (or Negative) Response
2.31 Expressing Difficulty
2.32 Inquiring About Difficulty
2.33 Expressing Ease
2.34 Inquiring About Ease
2.35 Expressing Belief/Opinion
2.36 Inquiring About Belief/Opinion
2.37 Forgetting
2.38 Comparing (Quality)
2.39 Remembering (Recalling, Reminding)
2.40 Rejecting (Fact, Situation)
2.41 Evaluating (Estimating, Assessing, Valuing, Judging)
2.42 Predicting
2.43 Hesitating
2.44 Describing/Narrating
2.45 Giving Examples/Citing
2.46 Classifying/Categorizing/Listing
2.47 Pointing Out Exceptions
2.48 Indicating Knowing/Ignorance
2.49 Trying/Proposing Solutions
2.50 Justifying/Presenting Excuses
2.51 Promising
2.52 Declaring/Stating
2.53 Protesting
2.54 Objecting/Resisting
2.55 Interviewing/Interrogating

F3 Expressing and Inquiring About Emotional Attitudes

3.1 Expressing Pleasure/Liking/Displeasure/Disliking
3.2 Inquiring About Pleasure
3.3 Expressing Satisfaction/Dissatisfaction
3.4 Inquiring About Satisfaction/Dissatisfaction
3.5 Expressing Disappointment

3.6 Expressing Fear/Worry
3.7 Asking About Fear/Worry
3.8 Expressing Surprise
3.9 Inquiring About Surprise
3.10 Stating Preference
3.11 Asking About Preference
3.12 Expressing Hope
3.13 Asking About Hope
3.14 Expressing Gratitude
3.15 Expressing Sympathy
3.16 Stating Want/Desire
3.17 Making an Emphatic Wish
3.18 Expressing Impatience
3.19 Indicating Quality of Performance
3.20.1 Setting Deadlines
3.20.2 Giving Reasons For Action/Non-Action
3.21 Inquiring About Impatience
3.22 Expressing Importance/Unimportance
3.23 Asking About Importance/Unimportance
3.24 Expressing Boredom
3.25 Expressing Happiness/Enthusiasm
3.26 Expressing Interest
3.27 Expressing Friendliness/Hostility
3.28 Expressing Trust/Suspicion
3.29 Expressing Admiration/Respect
3.30 Expressing Disrespect/Insults/Ridicule
3.31 Expressing Criticism/Blame/Accusation
3.32 Expressing Patience
3.33 Expressing/Inquiring About Complaint
3.34 Expressing Love/Hate

F4 Expressing and Determining Moral Attitudes

4.1 Apologizing
4.2 Expressing (Granting) Forgiveness
4.3 Expressing Approval/Disapproval
4.4 Asking About Approval/Disapproval
4.5 Stating Indignation
4.6 Stating Reproach
4.7 Expressing Indifference
4.8 Asking About Indifference
4.9 Expressing Embarrassment
4.10 Expressing Appreciation
4.11 Expressing Regret
4.12 Expressing Amazement

4.13 Expressing Amazement Ironically/Negative Undertone
4.14 Expressing Relief
4.15 Expressing Resignation
4.16 Expressing/Inquiring About Perplexity
4.17 Expressing That a Result Was Not Expected
4.18 Expressing Honor/Dishonor
4.19 Expressing Pride/Humility/Modesty
4.20 Expressing Moral/Religious Beliefs

F5 Expressing and Inquiring About Getting Things Done (Suasion)

5.1 Making Suggestions
5.2 Inquiring About Suggestions
5.3 Making Requests
5.4 Making (Expressing) Advice
5.5 Questioning Advice
5.6 Offering Invitation(s)
5.7 Giving Directions/Instructions/Commands
5.8 Making Threat(s)
5.9 Expressing Correction
5.10 Encouraging Someone to Perform
5.11 Inviting/Requesting Others To Perform An Action
5.12 Asking Someone To Hurry
5.13 Proposing a Course of Action
5.14 Persuading Someone to Do Something
5.15 Making/Changing Plans
5.16 Making/Expressing Decisions
5.17 Negotiating
5.18 Compromising
5.19 Asking if Someone Is Free/Busy
5.20 Making/Breaking/Avoiding Commitments
5.21 Manufacturing
5.22 Giving Orders

F6 Socializing (Engaging in Social Activities)

6.1 Greeting or Addressing (Subordinate, Peer, Superior), Forma and Informally
6.2 Taking Leave/Departing, Planning To Meet Again
6.3 Getting To Know Each Other (Sharing Likes/Dislikes, Exp ences, Ideas, Hobbies, Opinions)
6.4 Making General Statement Leading Up To Conversation
6.5 Terminating Conversation
6.6 Extending Wishes

6.7 Introducing People/Oneself/Someone
6.8 Responding To Introduction
6.9 Talking at the Dinner/Cafe/Restaurant Table
6.10 Proposing a Toast
6.11 Responding To a Toast
6.12 Striking a Bargain
6.13 Being Hospitable (Offering Food, Drinks, Etc.)
6.14 Presenting/Receiving Gifts
6.15 Asking/Offering Help
6.16 Telling Jokes/Anecdotes/Teasing
6.17 Expressing Thanks
6.18 Expressing Compliments, Congratulations, Praises, Flattery
6.19 Inquiring About Health/Welfare
6.20 Expressing Concern
6.21 Expressing Compassion
6.22 Pressuring Someone To Do Something
6.23 Acknowledging Polite Comment
6.24 Gossiping/Telling Secrets
6.25 Recounting Personal Experience/Boasting

7 *Managing Communication*

7.1 Interrupting/Acknowledging Interruptions
7.2 Sequencing Communication
7.3 Focusing on Topic
7.4 Refocusing and/or Adjusting Communication
7.5 Controlling/Tempo/Speed of Conversation
7.6 Controlling/Modulating Volume
7.7 Requesting Repetition or Offering To Repeat
7.8 Questioning
7.9 Requesting and/or Offering Translation, Explanation, or Clarification
7.10 Commenting on or Inquiring About Intelligibility
7.11 Commenting on a Topic/Subject
7.12 Changing/Returning To Topic/Subject
7.13 Passing on Information
7.14 Using Openers, Links, Responders
7.15 Obtaining/Insisting Upon Someone's Attention
7.16 Reporting Information Through the Media
7.17 Communicating Through Correspondence

8 *Telephone Behavior*

8.1 Answering a Telephone Call
8.2 Making a Telephone Call

8.4 Responding To Answer
8.5 Requesting To Speak To Someone
8.6 Responding To Request To Speak To Someone
8.7 Stating, "Wrong Number"
8.8 Putting Caller on Hold
8.9 Talking To Speaker
8.10 Returning a Telephone Call
8.11 Ending Telephone Conversation
8.12 Stating Reason For Call
8.13 Stating That One Does Not Have Any More Time To Talk

BEST COPY AVAILABLE

The CLEAR Oral Proficiency Exam (COPE)
Project Report Addendum: Clinical Testing and Validity and Dimensionality Studies

Lih-Shing Wang
Gina Richardson
Nancy Rhodes

(Addendum to *The CLEAR Oral Proficiency Exam (COPE) Project Report* by Shelley Gutstein and Sarah H. Goodwin, 1987)

This report was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract #400-85-1010. The opinions expressed in this report do not necessarily reflect the positions or policies of OERI or ED.

Center for Language Education and Research
Center for Applied Linguistics, Washington, D.C.
1988

Table of Contents

I. Introduction

II. Clinical testing of the COPE

III. Validity and Dimensionality of COPE

IV. Conclusion

Appendices

A. Revised COPE Cue Cards

B. Revised Rating Scale

C. Revised Instructions for Using COPE

ACKNOWLEDGEMENTS

We would especially like to thank Sarah Goodwin, Shelley Gutstein, Jane Gaytán, Lee Lundin, and Paquita Holland who helped enormously in the pilot testing phase of the project. Lynn Thompson, Karen Willetts, and Donna Christian also deserve much thanks for their substantial contributions to various phases of the project.

A special acknowledgement goes to Laurel Winston, John Karl, and Donna Sinclair who provided invaluable assistance in the final production of the test instrument and of this report.

Finally, we would like express our great appreciation to the students and teachers at the schools who volunteered to pilot our instrument.

I. Introduction

This report is a follow-up to the Project Report of *The CLEAR Oral Proficiency Exam (COPE)* , by Shelley Gutstein and Sarah H. Goodwin (1987). In this addendum we will describe the clinical testing of the COPE-Spanish with 36 fifth and sixth grade students in a partial immersion program in the midwest, 12 sixth grade students in a content-based FLES program in the same school district, and 65 fifth and sixth graders in a two-way immersion program on the east coast.

The validity testing of the COPE will be discussed in terms of a comparison with another Spanish oral proficiency test, the IDEA Oral Language Proficiency Test (IPT). The dimensionality of the COPE will be discussed in relation to whether the four subscales of the test -- comprehension, fluency, vocabulary, and grammar -- represent necessarily separate entities or whether they measure one single construct, i.e., general oral language proficiency.

II. Clinical Testing of the COPE

Testing locations and sample size. Three programs were selected for clinical-testing of the COPE. Because we were interested in testing students with as wide a range of language proficiency as possible, we selected three different types of programs. One was a partial immersion program in a school district in the midwest, the second was a content-based FLES program in the same district, and the third was a two-way partial immersion program on the east coast. The total sample was 113 students: 36 students at the partial immersion site, 12 at the content-based FLES site, and 65 at the two-way partial immersion site.

Site Descriptions

Partial Immersion Program. The Spanish partial immersion program is in a K-6 school located in a midwestern metropolitan school district. The program is located at a "basic skills" school that parents in a specific geographical region have the option of selecting for their children. The school environment consists of graded, self-contained classrooms with a highly structured curriculum emphasizing academic achievement in the basic skills. The partial immersion section of the school has been in existence for 3 years, since the fall of 1985. Three subjects, social studies, science, and math, are taught exclusively in Spanish in grades K-6. Before 1985 the students were receiving daily instruction in Spanish through a typical FLES approach. The ethnic make-up of the student body is: 47% White, 34% Black, 11% American Indian and Asian American, and 8% Hispanic. A small number of native Spanish speakers are included in the partial immersion program though it is not designed as a two-way program.

Some of the students tested had been in the program for its three years of existence and had also participated in the FLES classes offered previously. Other students were in their first year of the program, while others had been in the program three years but had not been there for the previous years of FLES. Native speakers were not included in the testing.

Content-based FLES Program. The content-based FLES program is located in a K-8 school in the same district as the partial immersion program. Students in grades K-6 are taught the social studies curriculum entirely in Spanish. The material covered is not repeated in English -- the only social studies instruction they get is in Spanish. Students in grades 7-8 are taught a more general language class which does include some science and social studies content. The fifth and sixth graders tested receive 55 minutes of daily instruction using this combined Spanish and social studies approach. Although the program itself has been in existence for a number of years, there is a high rate of turnover among students as in any urban school. A third of the students tested had just entered the program and were in their first year of Spanish while the rest had been in the program for four, five, or six years.

Two-way Partial Immersion Program. The two-way immersion school is a public elementary school in a metropolitan area that provides a bilingual education for students from pre-kindergarten through sixth grade. Intensive instruction is provided in Spanish and English, and both languages are used for presenting social studies, science, math, language arts/reading, and computer literacy. The language of instruction for social studies, science, and math varies according to periodic rotations. Approximately 50% of the day is devoted to instruction in each language. Every classroom has

two full-time teachers, one a native English speaker and one a native Spanish teacher. The classes are made up of native English speakers and native Spanish speakers; 60% of the student body is Hispanic. At this school both native and non-native Spanish speakers were tested.

Testing Procedures

CLEAR Oral Proficiency Exam (COPE). The COPE was administered at the two-way immersion school by two examiners simultaneously. The first examiner read the instruction cards to the examinees, prepared the appropriate sequence of cards in case deviation from the core cards was necessary, and provided subtle encouragement in sustaining the conversation (maintaining eye contact, perhaps saying "that's interesting!" before moving on to the next card, etc.). The second examiner was responsible for rating the students according to the COPE rating scale.

The test was administered in one of two rooms, depending on which was available. One room was a multi-purpose room used for meetings, music classes, and various types of special classes. Examiners and students sat around one end of a long table, with the examiner who was reading the cards sitting closest to the students. The disadvantage was that examinees were sometimes distracted by other activity in the room -- younger students passing through on their way to class, telephones ringing in a nearby teacher's office, etc. When this room was no longer available, testing was moved to the library. Students and examiners sat at a round table, with the students close enough to each other to preserve a natural conversational distance. This arrangement proved superior because the library was generally isolated from the normal activities of other students. A rug absorbed many distracting sounds.

Students tested in each setting were tape recorded so that 1) ratings could be clarified later if examiners felt that they were unable to achieve an accurate rating at the time; 2)students' performance could be reviewed in order to improve the accuracy of the rating scale; and 3) examiner's performance could be observed for its influence on student behavior. The tape recorder was placed between the two students, and their permission was obtained before any actual recording was done.

Pairs of students were selected by classroom teachers so as to be roughly matched for ability in Spanish. Pairs were summoned by examiners at approximately 25-minute intervals throughout the school day (20 minutes for testing and 5 for rating each pair). For the most part students were eager to participate and "show off" their Spanish to the two interested and sympathetic examiners. At the close of the pilot testing period, students were presented with certificates recognizing their participation in the project.

The administration of the test at the partial immersion and content-based FLES schools was similar to the administration at the two-way immersion school, except that one examiner performed both the administering and the rating functions. From discussions with the raters, it was found that it was difficult for one rater to fulfill both functions of administering the test and rating the students without extensive experience in administering the test. It was determined that using two examiners provided a distinct advantage over only one examiner. With two examiners, the first one can focus exclusively on administering the test while the other can focus entirely on rating the students.

IDEA Oral Language Proficiency Test (IPT-1). The IPT test was selected as the oral proficiency test to be administered at the same time as the COPE to assess its validity. The IPT was administered individually to the students at the three sites who were administered the COPE. For reliability purposes, we wanted to have a different rater for the COPE and the IPT. Therefore, the IPT test administrator was not the same person as the examiner for the COPE. This way we hoped to avoid artificially inflating the validity index -- a possible result of using the same rater who might assign a second rating based on knowledge of the student's performance on the previous test. The administration of the test took from 5-15 minutes, depending on the proficiency level of the student.

The IPT was designed to measure native Spanish speakers' oral proficiency in Spanish. The test consists of 83 items, with each item testing one of six oral language skill areas: syntax, morphology, lexicon, phonology, comprehension, and oral expression. During each administration, only one student is tested at a time. The student is required to respond to the questions presented either verbally or visually. The student advances until the test is completed or stops at a proficiency level as indicated by the number of errors committed at that level. Student performance is rated on a scale from A-F, with an additional possible category of M, which designates mastery of the test. The scale can then be collapsed into a three-category scale: NSS (Non-Spanish Speaking), LSS (Limited Spanish Speaking), and FSS (Fluent Spanish Speaking).

III. Validity and Dimensionality of COPE

Validity refers to the extent to which a test measures what it is intended to measure. Among the many types of test validity, concurrent validity, or the extent to which a test score corroborates the result of an independent external criterion measure administered at the same point in time, is examined here. For this study, the criterion measure against which the COPE is validated is the IDEA Oral Language Proficiency Test (IPT).

Validity of IPT. In order to validate the COPE against the IPT, it is important to understand how the IPT was originally validated. Validity of the IPT was assessed by the test authors in three categories: content validity, criterion-related validity, and construct validity (Enrique F. Dalton, *IPT Technical Manual. IDEA Oral Language Proficiency Test - Spanish.* Whittier, CA: 1980). Content validity was measured by the extent to which items on the IPT assess the six skill domains that the authors consider pertinent to oral language proficiency . These "domains" are syntax, morphology, lexicon, phonology, comprehension, and oral expression. Each item was analyzed to determine which of these domains is tapped, and the analysis was used to construct a test blueprint. Given the fact that these domains mix language components (syntax, morphology, lexicon, and phonology) with language skills (comprehension and oral expression), all items were found to assess more than one domain.

All items (100%) measured comprehension since the responses were elicited through questions or oral instructions. Similarly, 90% of the items measured oral expression, since they required a verbal response as opposed to pointing or an action (such as standing) on the part of the examinee. It

was possible for other items to assess more than one component of language, e.g., a past tense /t/ morpheme would assess both phonology and morphology. Using this type of analysis, it was determined that the components of language were assessed by the following percentages of items: syntax - 45%, lexicon - 88%. phonology - 69%, and morphology - 35%. Given the above findings, the IPT authors concluded that the six content domains are adequately sampled by the 83 items on the IPT.

Criterion-related validity of the IPT was examined through a study that correlated teachers' predicted IPT level classifications with actual IPT classifications. The obtained correlation was .79 (N=1122). In two additional studies, IPT scores were converted to Fluent English Speaking (FES), Limited English Speaking (LES), and Non-English Speaking (NES) classifications. In one study the classifications were correlated with the FES/LES/NES classifications obtained using five other tests approved by the California State Department of Education. A correlation of .75 (N=721) was found between the classification obtained using the IPT and the classification obtained using the other instruments.

In the final study, the IPT classifications were compared with FES/LES/NES classifications made by teachers on the basis of their knowledge of the student's level of oral language ability, academic ability, and other unobstrusive measures. A correlation of .71 was found (N=1200). The results of this and the two previously mentioned studies permitted the IPT authors to conclude that the IPT is a valid instrument for assessing oral language proficiency.

Validity of COPE. Operationally, the concurrent validity index of the COPE was measured by the Pearson product-moment correlation between the

total COPE and the IPT. The total COPE was coded as the sum of the four subscores -- comprehension, fluency, vocabulary, and grammar. Each subscore ranges from 1 (junior novice low) to 9 (superior). The IPT was coded on a scale of 1-7, representing the original A-F scale plus M.

As reported in Table 1, the Pearson product-moment correlation between the total COPE score and the IPT score for the total sample (N=113) is .62 ($p<.0001$). Although this is somewhat less than the conventional validity index criterion level (i.e., .75), the correlation is considered "reasonably good" because the IPT is very different from the COPE in many respects (e.g., format, content, context). In other words, the IPT may not be the "ideal" criterion which represents the most valid measure of the construct in question (i.e., oral proficiency), if one exists at all. Bearing this in mind, the fact that the COPE has a concurrent validity index of .62 should provide us with a fair degree of assurance that the COPE validly measures oral proficiency as intended.

When content-based FLES students (those who receive regular FLES instruction combined with a social studies class taught in Spanish) are considered separately from the partial immersion students, the validity index varies with group membership. For the content-based FLES students, the validity index is .81 ($N=12$, $p<.001$); whereas for the partial immersion students, the validity index is .57 ($N=101$, $p<.0001$). Because of the small sample size of the content-based FLES group, it is difficult to speculate the reason for the difference in the validity indices. A significance test on the difference between the two indices indicates that the difference is not significant at the .05 level ($z=1.40$). This suggests that the COPE may be an equally valid measure of oral proficiency for both content-based FLES and partial immersion students.

Table 1. Concurrent Validity of COPE.

Group	N	Validity index	p<
Content-based FLES	12	.81	.001
Partial immersion	101	.57	.0001
Total sample	113	.62	.0001

Dimensionality of COPE. The dimensionality of the COPE pertains to the question of whether the four subscales -- comprehension, fluency, vocabulary, grammar -- represent "psychologically real" entities that comprise general oral language proficiency and yet are empirically separable. Two analytical approaches were used to examine the dimensionality of the COPE: (a) intercorrelations among the four subscales, and (b) principal component analysis.

Intercorrelations among the four subscales, again measured by Pearson product-moment correlations, are reported in Table 2. The intercorrelation patterns clearly indicate that the four COPE skills -- comprehension, fluency, vocabulary, grammar -- are highly intercorrelated. The intercorrelations range from .82 to 1.0 for the content-based FLES students, from .95 to .99 for the partial immersion students, and from .95 to .99 for the total group. This suggests that the four oral proficiency skills measured by the COPE are essentially indistinguishable. In other words, the four COPE subscales measure one single underlying construct, i.e., general oral proficiency.

Table 2. Intercorrelations Among Four COPE Subscales.

	1	2	3	4
1. Comprehension	-			
2. Fluency	.88 .99 .99	-		
3. Vocabulary	.88 .97 .97	1.00 .98 .98	-	
4. Grammar	.82 .95 .95	.92 .96 .96	.97 .96 .97	-

Note: The first number in each cell refers to the content-based FLES group (N=12); the second the partial immersion group (N=101); the third the total group (N=113).

Further evidence arguing for the unidimensionality of the COPE is provided by the principal component analysis results reported in Table 3. The eigenvalues of the extracted principal components clearly indicate that only the first principal component is significant, explaining 94% of the total variance for the content-based FLES group, 98% for the partial immerison group, and 98% for the total group. Therefore, one single principal component is sufficient to explain the total variance of the four COPE skills. This common principal component can be safely labelled as "general oral language proficiency."

Table 3. Eigenvalues of the Principal Components Extracted from the Correlation Matrices.

Principal component	Group					
	Content-based FLES		Partial Immersion		Total	
	Eigen-value	Propor-tion	Eigen-value	Propor-tion	Eigen-value	Propor-tion
1st PC	3.76	.94	3.90	.98	3.91	.98
2nd PC	.21	.05	.06	.01	.05	.01
3rd PC	.03	.01	.03	.01	.03	.01
4th PC	.00	.00	.01	.00	.01	.00

IV. Conclusion

The recommendations presented in the original COPE project report concerning the Spanish version of the COPE have been addressed in this follow-up report. As suggested, a large sample (113 students) was used to test the validity and dimensionality of the test. The dialogue cards have now been revised to incorporate teachers' and test administrators' suggestions and re-ordered according to their level of difficulty. In addition, the dialogue cards were reviewed by a native speaker of Spanish to check for accuracy and appropriateness.

The results of the validity testing provided us with assurance that the COPE measures oral proficiency as intended. There was no significant difference when comparing validity for partial immersion and content-based FLES students, suggesting that the COPE may be an equally valid measure of oral proficiency for both types of programs. Although students in a total immersion program were not tested, it is hypothesized that that the COPE would be a valid measure for them as well.

Results of the dimensionality test suggest that the four subscales of proficiency skills measured by the COPE (comprehension, fluency, vocabulary, and grammar) actually measure one single underlying construct -- general oral proficiency. There was a very high correlation among the four skills. This means that in the future instead of giving each student four subscores on the test, all that is needed is a global score which can be an average of the four scores.

For further research on the COPE, it is suggested that the test be administered to *total* immersion students, to test the hypothesis that the instrument would be a valid measure for those students as well. In addition, further study should include additional administrations to content-based FLES students as well as include studies on inter-rater reliability.